

# Verification and Validation of Digital Pathology (Whole Slide Imaging) for Primary Histopathological Diagnosis: All Wales Experience

M. Babawale<sup>1</sup>, A. Gunavardhan<sup>1</sup>, J. Walker<sup>2</sup>, T. Corfield<sup>1</sup>, P. Huey<sup>1</sup>, A. Savage<sup>1</sup>, A. Bansal<sup>1</sup>, M. Atkinson<sup>1</sup>, H. Abdelsalam<sup>1</sup>, E. Raweily<sup>1</sup>, A. Christian<sup>3</sup>, I. Evangelou<sup>3</sup>, D. Thomas<sup>3</sup>, J. Shannon<sup>4</sup>, E. Youd<sup>4</sup>, P. Brumwell<sup>4</sup>, J. Harrison<sup>5</sup>, I. Thompson<sup>5</sup>, M. Rashid<sup>6</sup>, G. Leopold<sup>6</sup>, A. Finall<sup>6,7</sup>, S. Roberts<sup>8</sup>, D. Housa<sup>9</sup>, P. Nedeva<sup>9</sup>, A. Davies<sup>1</sup>, D. Fletcher<sup>1</sup>, Muhammad Aslam<sup>1</sup>

<sup>1</sup>Department of Cellular Pathology, Glan Clwyd Hospital, Betsi Cadwaladr University Health Board, Bodelwyddan, Wales, <sup>2</sup>Department of Anaesthetics, Gwynedd Hospital, Betsi Cadwaladr University Health Board, Bangor, Wales, <sup>3</sup>Department of Histopathology, University Hospital of Wales, Cardiff and Vale University Health Board, Cardiff, Wales, <sup>4</sup>Department of Histopathology, Royal Glamorgan Hospital, Cwm Taf University Health Board, Rhondda Cynon Taff, Wales, <sup>5</sup>Department of Histopathology, Royal Gwent Hospital, Aneurin Bevan University Health Board, Newport, Wales, <sup>6</sup>Department of Cellular Pathology, Morriston Hospital Swansea Bay University Health Board, <sup>7</sup>Abertawe Bro Morgannwg Health Board, Port Talbot, Wales, <sup>8</sup>Department of Cellular Pathology, Singleton Hospital, Abertawe Bro Morgannwg Health Board, Port Talbot, Wales, <sup>9</sup>Department of Histopathology, Glangwili Hospital, Hywel Dda University Health Board, Carmarthen, Wales

Submitted: 25-Jun-2020

Revised: 03-Jul-2020

Accepted: 01-Oct-2020

Published: 23-Jan-2021

## Abstract

**Aims:** The study is aimed to verify Aperio AT2 scanner for reporting on the digital pathology platform (DP) and to validate the cohort of pathologists in the interpretation of DP for routine diagnostic histopathological services in Wales, United Kingdom. **Materials, Methods and Results:** This was a large multicenter study involving seven hospitals across Wales and unique with 22 (largest number) pathologists participating. 7491 slides from 3001 cases were scanned on Leica Aperio AT2 scanner and reported on digital workstations with Leica software of e-slide manager. A senior pathology fellow compared DP reports with authorized reports on glass slide (GS). A panel of expert pathologists reviewed the discrepant cases under multiheader microscope to establish ground truth. 2745 out of 3001 (91%) cases showed complete concordance between DP and GS reports. Two hundred and fifty-six cases showed discrepancies in diagnosis, of which 170 (5.6%) were deemed of no clinical significance by the review panel. There were 86 (2.9%) clinically significant discrepancies in the diagnosis between DP and GS. The concordance was raised to 97.1% after discounting clinically insignificant discrepancies. Ground truth lay with DP in 28 out of 86 clinically significant discrepancies and with GS in 58 cases. Sensitivity of DP was 98.07% (confidence interval [CI] 97.57–98.56%); for GS was 99.07% (CI 98.72–99.41%). **Conclusions:** We concluded that Leica Aperio AT2 scanner produces adequate quality of images for routine histopathologic diagnosis. Pathologists were able to diagnose in DP with good concordance as with GS. **Strengths and Limitations of this Study:** Strengths of this study – This was a prospective blind study. Different pathologists reported digital and glass arms at different times giving an ambience of real-time reporting. There was standardized use of software and hardware across Wales. A strong managerial support from efficiency through the technology group was a key factor for the implementation of the study. **Limitations:** This study did not include Cytopathology and *in situ* hybridization slides. Difficulty in achieving surgical pathology practise standardization across the whole country contributed to intra-observer variations.

**Keywords:** Concordance, diagnosis, digital pathology, glass slides, ground truth, light microscopy, scanned slides/whole slide imaging, validation, variances/discrepancies, verification

## INTRODUCTION

Histopathology is at the dawn of a new era in digital pathology (DP), with the wide availability of whole slide imaging (WSI), digital image analysis, electronic specimen labeling, tracking, and incorporation of laboratory management systems with digital dictation tools. DP tools, which have until now been

**Address for correspondence:** Dr. Muhammad Aslam,

Department of Cellular Pathology, Glan Clwyd Hospital, Betsi Cadwaladr University Health Board, Rhuddland Road, Bodelwyddan, LL18 5UJ, Wales.  
E-mail: muhammad.aslam3@wales.nhs.uk

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** reprints@medknow.com

**How to cite this article:** Babawale M, Gunavardhan A, Walker J, Corfield T, Huey P, Savage A, *et al.* Verification and validation of digital pathology (whole slide imaging) for primary histopathological diagnosis: All wales experience. *J Pathol Inform* 2021;12:4.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2021/12/1/4/307702>

Access this article online

Quick Response Code:



Website:  
[www.jpathinformatics.org](http://www.jpathinformatics.org)

DOI:  
10.4103/jpi.jpi\_55\_20

widely used in teaching and research, are beginning to make inroads into routine diagnostic histopathology reporting in Europe and Canada, and most recently in the United Kingdom (UK).<sup>[1,2]</sup> This technology offers opportunities of remote reporting for intraoperative frozen sections,<sup>[3]</sup> flexible working from home,<sup>[4]</sup> improved turnaround times for case referrals, and ready availability of archived surgical materials for multi-disciplinary team meetings (MDTM).

The recent developments in DP in Europe have closely followed the granting of European Conformity (CE) marking to some of the leading vendors of WSI scanners worldwide, such as Leica Biosystems.<sup>[5]</sup> Leica Biosystems obtained CE-IVD marking for the use of their Aperio AT2 scanner for diagnostic histopathology work in Europe. The Welsh Assembly in Cardiff recognized the potential of DP for improving the histopathology services, in turn positively influencing patient care. The Welsh Assembly approved efficiency through Technology Funding to carry out feasibility studies on the verification of Leica Aperio AT2 scanner and validation of pathologists to use DP for diagnostic purposes.

A review of existing English literature revealed origins and progress of DP in the last four decades. DP emerged in the 1980s with the development of telepathology technology. It was mostly used for second opinions and research purposes. WSI was first developed in 1990s. It refers to scanning a complete microscope slide, capturing multiple high-resolution images, and creating a full image of the histopathology section and saved as a single high-resolution digital file.<sup>[6]</sup>

There are various studies published in English literature on the validation of DP and concordance between digital and glass diagnosis. However, the case numbers in these studies were low.<sup>[7-10]</sup> A systematic review by Goacher *et al.* in 2016 highlighted a gap in the literature regarding evidence to validate the use of WSI in routine primary diagnosis.<sup>[2]</sup>

The introduction of any new technology needs validation and verification to ensure quality and patient safety. College of American Pathologists recommends validation studies before WSI systems to be used for diagnostic purposes.<sup>[11]</sup>

A validation study of 3017 cases by Snead *et al.* in 2016 and multicenter blinded noninferiority study of 1992 cases by Mukhopadhyay *et al.* in 2018 were the largest studies that established noninferiority of DP in literature at the time of writing this report. Both studies followed a wash out period between interpreting glass slides (GSs) and DP to reduce the intraobserver variation.<sup>[1,12]</sup>

All pathologists involved in this study participated in various regional and national level external quality assurance schemes as part of quality assurance and appraisal. All pathologists work in a general histopathology reporting setting, with specialist interest in one or two sub specialties. Multi header slide review meetings and MDT slide reviews are various modalities in place to reduce error and intraobserver variation.

The geographical variability of Wales with shortage of pathologists poses challenges in delivering pathology services. This study on validation and verification of DP is the first step of the digitization of pathology laboratories in Wales to improve patient care.

## MATERIALS AND METHODS

### Ethics

The study proposal was approved by the Clinical Audit and Effectiveness Department for approval to cover the review process from April 2016 to November 2016 as a national project. There was also an agreement developed between all the Welsh health boards to share information called “Intra NHS Information Sharing Agreement v1.0.”

### Case selection, slide scanning, and viewing

The study included representations from all tissue types received in Glan Clwyd Hospital (the study hub) over the study period (April – December 2016). The pathologists at Glan Clwyd Hospital took part in a pilot study before the commencement of the project, in line with recommended guidelines.<sup>[13]</sup> The remaining participating pathologists across the other seven Cellular Pathology departments across Wales had access to three test sets (20 cases per test set; 60 cases in all), aiming at improving their knowledge and developing their confidence in the use of DP. Other than that none of the pathologists had any previous experience in reporting histopathology slides in the digital platform. At the time of this study, there were no published guidelines by the Royal College of pathologists.

The workflow and protocol for slide scanning and reporting are shown in Chart 1. Leica aperio AT2 scanner located at Glan Clwyd hospital (Betsi Cadwaladr University Health Board) was the hub of this study. The scanned slides were allocated electronically to 22 out of 76 pathologists across the six Health boards in Wales [Table 1].

All the cases were scanned on the Aperio AT2 scanner at ×20 (0.5 um/pixel). After reviewing the studies available at that time, the efficiency through technology group who was

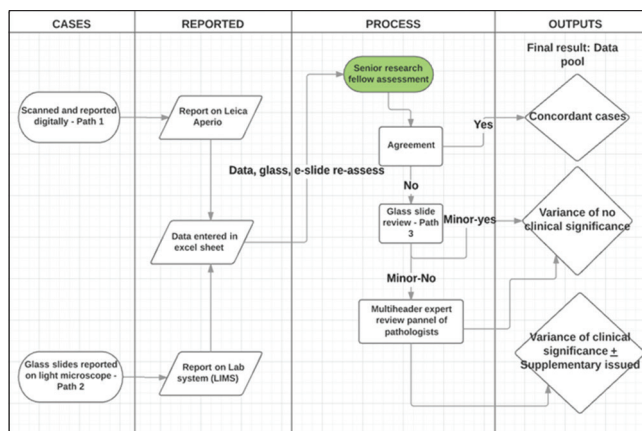


Chart 1: Process flow mapping

**Table 1: Cases reported by pathologists in various health boards**

Welsh health boards	Pathologists	Number of cases reported	Total for health board
Swansea Bay University Health Board	1	96	357
Aneurin Bevan University Health Board	2	134	
	3	127	
	4	98	291
	5	105	
	6	88	
Betsi Cadwaldr University Health Board (Hub of the study)	7	126	1146
	8	48	
	9	105	
	10	235	
	11	362	
	12	261	
	13	9	
Cwm Taf Morgannwg University Health Board	14	164	359
	15	133	
	16	62	
Cardiff and Vale University Health Board	17	168	503
	18	164	
	19	171	
Hywel Dda University Health Board	20	129	345
	21	128	
	22	88	
		Total	3001

key player in organizing the study, decided unanimously to scan at  $\times 20$  for the study project to have quick scanning and see the capabilities of the technique on that magnification. If needed, later on for clinical reporting, it would be easy to move upward on higher magnification but difficult other way around. The study used 1-inch  $\times$  3-inch (2.54 cm  $\times$  7.62 cm) microscope slides with 0.17 mm coverslip scanned automatically using the Autoloader. Line scanning method was used using a  $\times 20/0.75$  NA Plan Apo objective lens with automatic tissue finding and tissue focusing abilities and scanning at a resolution of 50,000 pixels per inch (0.5 m per pixel).

All participating pathologists were provided with workstations consisting of a 3.60 GHz processor (Intel® Core™ I7-4790 CPU, 16GB RAM, 64-bit operating system (Windows 7) with double monitors of high-resolution display with 1920  $\times$  1200 resolution and a refresh rate of 60 Hz.

24-bit color contiguous pyramid TIFF was immediately viewable using LCD (flat panel) monitors with 24" screen size supplied to all the pathologists with a color depth of 24 bit, the brightness of 400 cd/m<sup>2</sup> and contrast ratio of at least 1000/1. The images were viewable at  $\times 4$  up to  $\times 40$ . The eSlide file format was standard pyramid tiled TIFF with JPEG2000 image compression.

The images are analyzed using imagescope (version 12.0.1.5027) and slides were arranged using Aperio e-slide manager version 12.2.1.5005 (Leica Biosystems).

Cases were scanned prospectively. Each participant gets image of request cards and digital slides of the case. Those specialties that pathologists do not routinely report were not allocated to them. The reports were typed and saved in Leica software. The corresponding GSs were allocated to a different pathologist for live reporting and authorization. This study design was to eliminate recall bias and therefore washout periods were not necessary. Pathologists can request further work if needed, including extra levels, special stains, and immunohistochemistry. These slides were scanned digitally and allocated for the analysis. They can request coordinator for previous biopsy results if required.

### Establishment of concordance, variances, and discrepancies in diagnosis and ground truth

The independent observer compared the GS report with the DP reports. The discordant cases were reviewed by a third Pathologist, who has sub-specialty interest in the particular case. The GT was determined by reviewing the DP images and GSs. More difficult cases where GT was difficult to establish, the GSs were reviewed under multiheader microscope by a panel of pathologists. The panel's decision was accepted GT.

### Sample size calculation

Previous work by Snead *et al.*<sup>[1]</sup> demonstrated a concordance of 99.3% between DP and GS, and we adopted their approach that concordance of  $<98\%$  would be of concern. We calculated that a sample size of 3000 cases would have a statistical power of 99% to demonstrate such a difference at the 5% significance level. One of the risks with this kind of study is a type 2 statistical error when we conclude that the techniques are equivalent when in fact, they are not. The high degree of power makes such an error less likely.

### Statistical analysis

The cases with the agreement in diagnosis between DP and GS reports were called concordant. The discordant cases had inter-observer variances between DP and GS. They were further classified into discordant cases of no clinical significance (IVNCS) and discordant cases of clinical significance. The INVCS were benign lesions of no impact in patient management. The cases were classified as clinically significant was based on accepted criteria.<sup>[13]</sup> Examples include discordance between benign and malignant, grading of dysplasia, and missing findings that change patient management. Percentages of concordance (sensitivity) were calculated with 95% confidence intervals (CIs). We adopted the same criteria for non-inferiority as Snead *et al.*,<sup>[1]</sup> that non-inferiority would be demonstrated if the 95% CI remained  $>98\%$  concordance. In addition, a Bayesian binomial test was performed against the alternative hypothesis that the percentage of concordance was  $<98\%$ . All calculations were performed using Microsoft Excel and Jeffreys's Amazing Statistics Program (JASP) 2017.

## RESULTS

7491 slides were scanned for the 3001 cases studied, which resulted in a digital archive of 3.8 Terabytes [Table 2]. The number of cases scanned and reported in each subspecialty is shown in Table 3. In 35 cases, due to operational issues, same pathologist reported glass and digital slides but at different times. 100% concordance was achieved in those cases.

2745 out of 3001 cases showed complete concordance between GS and DP reports (91.4%, CI 90.4%–92.4%). 256 cases (8.5%, CI 7.6%–9.6%) showed variance between GS and DP reports, of these, 86 (2.9%, CI 2.3%–3.5%) were considered to be clinically significant.

Of the total discrepancies observed, 161 cases showed inter-observer variances of no clinical significance (IVNCS). In addition, quality control (QC) issues were responsible for the initial non-concordance in eight cases.

Seven out of the eight QC issues found in the study were either due to slides being scanned and not reported, inadequate number of slides scanned in some cases, bubbles interfering with reporting on DP or an incorrect report being entered onto the DP datasheet. All these were corrected, and concordance was achieved by re-scanning and re-allocating to the pathologist. Two cases of *Helicobacter pylori* were initially missed on low power scanned DP and corrected by re scanning at ×60 objective.

Biopsies and resections	3001
Total number of slides	7491
Total data generated	3.8 TB
Slides per case	
Minimum	1
Maximum	6
Mean	2.5
Range of data per slide	
Minimum	67 KB
Maximum	8.3 GB
Mean	500 MB

Subspecialty	Number of cases (percentage of total)	Concordance (%)	Discordant cases (percentage of subspecialty)	Clinically significant discordance (percentage of subspecialty)
Skin	1356 (45.2)	1298 (95)	98 (7.2)	38 (2.8)
Gastrointestinal tract	830 (27.7)	774 (93.2)	56 (6.8)	26 (3.1)
Gynaecology	463 (15.4)	397 (85.7)	66 (14.3)	15 (3.2)
Urology	140 (4.7)	110 (78.5)	30 (21.4)	7 (5.0)
Head and neck	85 (2.8)	82 (97.6)	2 (2.4)	0 (0.0)
Soft tissues	67 (2.2)	66 (98.5)	1 (1.5)	0 (0.0)
Breast	47 (1.6)	45 (95.7)	2 (4.3)	0 (0.0)
Other general	13 (0.4)	12 (92.3)	1 (9.1)	0 (0.0)

On final analysis, there were 171 nonclinically significant discordant cases, 2915 concordant cases, and 86 clinically relevant discrepancies.

Of the 86 clinically significant variances, the Ground Truth lay with DP in 28 and with GS in 58 cases. In total, ground truth lay with DP for 2943 cases, giving a sensitivity of 98.07% (CI 97.57%–98.56%; Bayes factor 350); and with GS for 2973 cases, giving a sensitivity of 99.07% (CI 98.72%–99.41% Bayes factor 1641). Although a high degree of concordance was demonstrated, we were unable to meet the noninferiority criteria of >98% concordance.

Bayes factors are a way of quantifying evidence for or against the hypotheses being tested. A Bayes factor of 2 means that a particular hypothesis is twice as likely; whereas a Bayes Factor of 0.5 would indicate that a particular hypothesis is half as likely. A Bayes factor of 1 indicates no evidence in either direction.

For the Bayesian calculations, we set the null hypothesis, as “sensitivity is 98% or better,” and measured the evidence for this. In the case of DP, the Bayes factor of 350 means that the null hypothesis (sensitivity 98% or better) is 350 times more likely than the alternate hypothesis (sensitivity worse than 98%). This is interpreted as very strong evidence for the null hypothesis.

### Data analysis of dermatopathology cases

A total of 1356 dermatology samples were studied, in which concordance was achieved in 1298 cases (95%). There were 98 (7.2%) discordant cases, of which clinically significant ones were 38 (2.8%)

The main discrepancies were in diagnosing dysplasia and categorizing melanocytic lesions. One case of fungi was not identified on DP. In all cases except one, ground truth lay with GS [Table 4].

### Data analysis of gastrointestinal tract cases

In this study, there were 830 samples from the gastrointestinal tract (GIT). Concordance was achieved in 774 (93.2%) cases. 56 (6.8%) cases showed discordance between DP and GS of which 26 (3.1%) were deemed clinically significant.

Appropriate categorization of dysplasia was problematic in 11 cases, ranging from over calling to under calling of dysplasia. One case had difficulty in distinguishing high-grade

**Table 4: Dermatopathology discrepant cases**

Subspecialty	GS diagnosis	DP diagnosis	Third pathologist/multiheader diagnosis	GT	Number of repeats
Skin	Lichenoid inflammation and moderate dysplasia	Lichenoid inflammation. No dysplasia seen	Lichenoid inflammation and moderate dysplasia	GS	Undercall of dysplasia on DP, ×12
Skin	Compound melanocytic naevus	Skin showing dysplastic naevus	Compound melanocytic naevus	GS	Overcall of dysplasia on DP, ×4
Skin	Dysplastic junctional naevus	Lentigo maligna <i>in situ</i>	Dysplastic junctional naevus	GS	None
Skin	Lentigo maligna with early microinvasion	Junctional melanocytic naevus	Lentigo maligna with early microinvasion	GS	None
Skin	Malignant Melanoma in situ in radial growth phase	Severely dysplastic naevus. No evidence of melanoma	Malignant melanoma in situ in radial growth phase	GS	None
Skin	Spitz naevus	Malignant melanoma	Spitz naevus	GS	None
Skin	Actinic keratosis with moderate squamous atypia	Inflamed seborrheic keratosis	Actinic keratosis with moderate squamous atypia	GS	×3
Skin	Irritated seborrheic keratosis	Hypertrophic actinic keratosis	Hypertrophic actinic keratosis	DP	×2
Skin	Skin biopsy showing fungal infection and reactive atypia	Chronic nonspecific dermatitis. There is no atypia or malignancy	Pseudo-epitheliomatous hyperplasia secondary to fungal infection	GS	None
Skin	A and B: BCC, completely excised	A: BCC, completely excised B: Scanned but not reported	A: BCC, completely excised B: Scanned but not reported	QC issues	Similar QC issues on DP, ×7

GS: Glass slide, DP: Digital pathology, GT: Ground truth, BCC: Basal cell carcinoma, QC: Quality control

dysplasia from early invasive carcinoma. Two cases of hyperplastic polyps were called sessile serrated lesion (SSL) in DP. In all these cases, ground truth lay with GS.

In one case, intraepithelial lymphocytes (IELs) appeared more prominent, and pathologist reporting DP raised the possibility of Lymphocytic colitis. It was reported as normal large bowel mucosa on GS. Conversely, a case of increased IEL in small bowel mucosa in an early case of the coeliac disease diagnosed on GS was not identified on DP.

Two cases of *H. pylori*-associated active chronic gastritis correctly diagnosed on DP (viewed at x 40 magnification), were not identified on GS. In addition, two cases of nonspecific chronic gastritis on GS (also viewed as x 40 magnification) were reported as *H. pylori*-associated active chronic gastritis on DP. In fact, they were artifacts on the gastric mucosa mistaken for *H. pylori* like organisms [Table 5].

#### Data analysis of lymphoid cases

Two discrepant cases of lymphoid lesions were reported. A gastric biopsy reported as chronic gastritis in GS and ground truth was Marginal Zone Lymphoma on DP (reported by a pathologist with interest in lymphomas). In the second case, lymphocytes in a nasal polyp was interpreted as a Non-Blastic, T-Cell Non-Hodgkin's Lymphoma on DP (reported by a general pathologist), which was confirmed as a reactive lymphoid condition on GS.

#### Data analysis of gynecological cases

463 gynecologic samples were studied. Concordance was achieved in 397 (85.7%) cases. 66 (14.3%) cases were discrepant, of which 15 (3.2%) were clinically significant.

Pathologists found difficulty in categorizing dysplasia with DP in nine cases. Ground truth lay with GS in nine and with

DP in one case. Immunohistochemistry with p16 and ki67 was performed to finalize diagnosis during multi header panel review. One case of chorioamnionitis and one case of endometriosis were not identified in DP [Table 6].

#### Data analysis of urology cases

One hundred and forty urology samples were included in this study. 110 (78.5%) cases showed concordance. 30 (21.4%) cases showed discordance between DP and GS. Seven of the discordant cases were clinically significant (5.0%).

In prostatic biopsies, five cases of prostatic adenocarcinoma Gleason score 3 + 3 = 6 diagnosed on GS were incorrectly diagnosed as either benign prostate, prostatic adenosis, atypical acinar proliferation or post atrophic hyperplasia with focal prostatic intraepithelial neoplasia (PIN) on DP.

In one bladder biopsy, focal Grade 3 transitional cell carcinoma and focal carcinoma *in situ*, correctly reported on GS were missed on DP [Table 7].

There was no clinically significant discordance recorded for head and neck, soft tissue, breast and other general histopathology samples.

Reporting pathologists found the digital platform ergonomically friendly. No one experienced eye strain or back pains throughout the study. One pathologist reported an exacerbation of neck pains due to an old injury and another pathologist reported hand/wrist pain while navigating with mouse.

#### DISCUSSION

This study is a large multicenter study by 22 pathologists in the UK on DP to verify the use of DP for routine diagnostic

**Table 5: Clinically significant discrepancies in gastrointestinal tract**

Sub speciality	Glass diagnosis	Digital diagnosis	3 <sup>rd</sup> pathologist/multi header review diagnosis	Ground truth	Number of repeats
GIT	Hyperplastic polyp	Tubular adenoma with low grade dysplasia	Hyperplastic polyp	GS	Overcall of dysplasia on DP, ×6
GIT	Barrett's oesophagus with low grade dysplasia	Barrett's oesophagus No dysplasia seen	Barrett's oesophagus with low grade dysplasia	GS	Undercall of dysplasia on DP, ×4
GIT	TVA with high grade dysplasia	Moderately differentiated invasive adenocarcinoma in a TVA with high grade dysplasia	Moderately differentiated invasive adenocarcinoma in a TVA with high grade dysplasia	DP	None
GIT	Hyperplastic polyp	Sessile serrated adenoma with low grade dysplasia	Hyperplastic polyp	GS	×2
GIT	Normal large bowel mucosa	Large bowel mucosa with features of lymphocytic colitis	Normal large bowel mucosa	GS	None
GIT	Duodenal mucosa with features of early coeliac disease	Normal duodenal mucosa	Duodenal mucosa with features of early Coeliac disease	GS	None
GIT	Nonspecific chronic gastritis	<i>H. pylori</i> active chronic gastritis	<i>H. pylori</i> active chronic gastritis	DP	×2
GIT	Nonspecific chronic gastritis	<i>H. pylori</i> associated chronic gastritis	Nonspecific chronic gastritis	GS	×2
GIT	<i>H. pylori</i> associated chronic gastritis	Gastric mucosa suggestive of marginal zone lymphoma of MALT type	Gastric mucosa suggestive of marginal zone lymphoma of MALT type	DP	None
GIT	Normal small bowel mucosa	Sections of irritated seborrhoic keratosis	Pathologist reporting DP wrongly entered a different report on data sheet	QC issues	Misidentification error, ×1

*H. pylori*: *Helicobacter pylori*, TVA: Tubulo villous adenoma, DP: Digital pathology, GIT: Gastrointestinal tract, GS: Glass slide, MALT: Mucosa-associated lymphoid tissue

**Table 6: Discrepancy in gynaecologic pathology cases**

Glass diagnosis	Digital diagnosis	3 <sup>rd</sup> pathologist/multi header review diagnosis	Ground truth	Number of repeats
CIN 2, HPV changes and cervicitis	Cervicitis. No evidence of CIN seen	CIN 2 and HPV changes and cervicitis	GS	Undercall of CIN 2 on DP, ×5
HPV changes and CIN 1	HPV changes and cervical epithelia changes amounting to CIN 2	HPV changes and CIN 1	GS	Overcall of CIN 2 on DP, ×3
CIN 2 with cervicitis	CIN 1	CIN 1 (upheld by p16 and Ki 67 IHC)	DP	None
2 <sup>nd</sup> trimester placental chorioamnionitis	Normal 2 <sup>nd</sup> trimester placenta	2 <sup>nd</sup> trimester placental chorioamnionitis	GS	None
Chronic salpingitis and endometritis	Chronic salpingitis. No evidence of endometritis	Chronic salpingitis and endometritis	GS	None

DP: Digital pathology, GIT: Gastrointestinal tract, GS: Glass slide, CIN: Cervical intraepithelial neoplasia, IHC: Immunohistochemistry, HPV: Human papillomavirus

histological reporting and to validate the participating pathologists in reporting on DP.

Of the 3001 cases, DP was demonstrated to be correct in 98.1% (CI 97.6%–98.6%), which is similar to the seminal study of Snead *et al.*, 2016.<sup>[1]</sup>

Major discrepancy between DP and glass diagnosis in our study was 2.9%. This is comparable to the discrepancy rates in other similar studies in the literature.<sup>[1,12,30]</sup> Table 8 compares discrepancy rates between glass and digital as well as specialty wise categorization in other similar studies.

The existing average discrepancy rate in glass diagnosis calculated from Welsh External Quality Assurance Scheme and

Glan Clwyd hospital intra department audit is 3%. According to Raab *et al.* who reviewed self-reported discrepancies from 75 laboratories participating in the College of American Pathologists Q-Probe program, the overall discrepancy rate was about 7%, and major discrepancy was 5.3%.<sup>[16]</sup> Similarly, a retrospective review of 715 s - opinion slide reviews at Sun Yat-Sen Cancer Centre showed a major discrepancy rate of 6%.<sup>[17,18]</sup>

The 97.6%–98.6% CI obtained for the DP cases showed that we were not able to demonstrate noninferiority to the 98% limit that was originally specified. The decision to use Bayes' Factors as well as traditional frequentist statistics was originally taken to provide additional evidence had noninferiority been

**Table 7: Discrepancies in uropathology**

Subspecialty	Glass diagnosis	Digital diagnosis	Multiheader review diagnosis	Ground truth	Number of repeats
Urology	Prostatic adenocarcinoma Gleason 6 (3+3)	Focal prostatic adenosis	Prostatic adenocarcinoma Gleason 6 (3+3)	GS	Focal adenocarcinoma called benign on DP, ×5
Urology	Bladder mucosa with focal CIS and suspicious focus of G3 TCC	Ulcerated bladder mucosa with acute on chronic inflammation	Bladder mucosa with focal CIS and suspicious focus of G3 TCC	GS	None

CIS: Carcinoma *in situ*, TCC: Transitional cell carcinoma, GS: Glass slide, DP: Digital pathology

**Table 8: Comparison of discrepancies with similar studies in literature**

	Snead <i>et al.</i> 2016 <sup>[1]</sup>	Mukhopadhyay <i>et al.</i> 2017 <sup>[11]</sup>	Borowsky <i>et al.</i> 2020 <sup>[28]</sup>	Current study
Overall major discrepancy among pathologists using glass slides (%)	0.4	4.6	3.2	--*
Overall major discrepancy between WSI and signed out report (GT) (%)	0.7	4.9	3.6	2.9
Major discrepancy between WSI and signed out report (GT) in GIT (%)	0.4	<1	3.3	3.1
Major discrepancy between WSI and signed out report (GT) in gynaecological pathology (%)	0.26	3.1	3.18	3.2
Major discrepancy between WSI and signed out report (GT) in dermatopathology (%)	0.7	4.9	4.74	2.8
Major discrepancy between WSI and signed out report (GT) in uropathology (%)	2.4	5	3	5

\*3%: Average of existing discrepancy rate from Welsh external quality assurance scheme and Glan Clwyd Hospital intra departmental audit. WSI: Whole slide imaging, GIT: Gastrointestinal tract, GT: Ground truth

shown. It is interesting that the Bayes analysis shows very strong evidence in favor of the null hypothesis of no difference between techniques.

Our data adds more evidence to the existing literature about the concordance of digital image reporting with GS for routine diagnostic work.

Most of the validation studies published to date have highlighted a tendency to overcall dysplasia in most organ systems of the body, such as in resections and biopsies from the GIT, skin, and cervix.<sup>[1,19]</sup> Our study also showed a similar picture with discordance in diagnosing dysplasia with DP.

In our study, diagnostic discordance of clinical significance was observed in various melanocytic lesions. A study by the Dutch group also mentions about challenges to report on DP.<sup>[18]</sup> The wide variability in pathologist's diagnosis of invasive melanoma and melanocytic proliferations was buttressed by a recent study that examined 240 skin biopsies by 1187 pathologists, found that diagnoses within the disease spectrum from moderately dysplastic nevi to early-stage invasive melanoma were neither accurate nor reproducible.<sup>[21]</sup>

There were two cases correctly reported as actinic keratosis on DP, which were incorrectly reported as seborrheic keratosis on GS. Conversely, three cases of actinic keratosis correctly reported on GS were incorrectly reported as seborrheic keratosis or inflammatory changes on DP. Further investigations revealed that the errors were due to failure to adhere to the diagnostic criteria of clinical entities, irrespective of whether pathologists were reporting on the DP or GS.

The fact that fungal spores were missed on DP in one case of epitheliomatous hyperplasia secondary to fungal infection

correctly diagnosed on GS highlights the importance of requesting ancillary investigations.

We suggest that the threefold increase in under calling dysplasia on DP in our study when reporting skin specimen might be due to pathologists being aware of the "phenomenon of overcalling dysplasia on DP," thereby exercising greater caution in estimating dysplasia in epithelia on DP and in the process, overcompensated in their assessment by under calling dysplasia on DP. Given time and further exposure to reporting on DP, we believe that adjustments would be made by users to find a middle ground, where the accurate diagnosis of dysplasia on DP would become routine.

A tendency to overcall dysplasia in Barrett's esophagus in DP has been previously reported.<sup>[1,22]</sup> However, other groups have also reported the reverse, that a tendency to downgrade dysplasia when reporting GIT biopsies on digital microscopy.<sup>[23]</sup> In GI biopsies and resections, we found four cases with overcall of dysplasia on DP and another four cases with undercall. This shows a significant learning curve for dysplasia diagnosis in DP as described in the systemic analysis of discordant diagnosis by Williams *et al.*<sup>[22]</sup>

Interestingly in our study, two cases of *H. pylori*-associated active chronic gastritis were correctly diagnosed on DP at ×40, but missed on GS. Conversely, two cases of non-specific chronic gastritis, correctly reported on GS were incorrectly reported as *H. pylori*-associated active chronic gastritis on DP at x 40 magnification, as artifacts on the gastric mucosa were mistaken for *H. pylori* like organisms. One reason for a higher pickup rate of *H. pylori* like organisms associated active chronic gastritis in our study might be the awareness by our pathologists of a tendency to

miss the organisms on the DP viewing platform; hence more caution was exercised when reporting gastric biopsies with an active chronic inflammatory cell background. The accurate diagnosis of the two cases of *H. pylori*-associated active chronic gastritis at  $\times 40$  magnification might at best have been intuitive rather than a true ability to view the organisms clearly at  $\times 40$  on DP, as the majority of our pathologists admitted having difficulties viewing the organisms clearly at  $\times 40$ . As previously suggested,<sup>[1]</sup> scanning at  $\times 60$  for cases suspected of having bacteria or other organisms would be prudent in all such cases.

In the two cases of colonic hyperplastic polyps correctly diagnosed on GS, that were incorrectly called SSLs on DP; on review by our panel of pathologists, such errors were attributed to lack of adherence to diagnostic criteria on the part of the DP reporting pathologists, rather than inherent problems with the images on the DP platform.

In the literature, there are mixed reports on the accuracy of reporting prostatic biopsies on DP/WSI. In one study, there was a poor degree of consensus in the evaluation of the inter-observer variability in the assessment of atypical foci in twenty WSI of prostate biopsies among experts (5 experts and 7 nonspecialists).<sup>[24]</sup> Another study on fifty prostatic biopsies conducted by four expert urologic pathologists, found that reporting prostatic cancer parameters on WSI/VM were comparable to reporting on routine light microscope.<sup>[25]</sup> In our study, reporting prostatic biopsies was equally challenging. Six cases of prostatic adenocarcinoma, Gleason score  $3 + 3 = 6$ , correctly diagnosed on GS were missed on DP and incorrectly diagnosed either as benign prostatic tissues, prostatic adenosis, atypical acinar proliferation or post atrophic hyperplasia with the focal PIN. We found that three of the prostatic adenocarcinoma cases, Gleason  $3 + 3 = 6$  reported as benign on DP were by a general pathologist who did not routinely report prostate biopsies. In the remaining three cases, the lesional areas in the prostatic cores were identified but incorrectly interpreted as either atypical acinar proliferation, prostatic adenosis, or hyperplasia with focal PIN on DP. A number of the participating pathologists at Glan Clwyd Hospital expressed concern regarding reporting prostate biopsies on DP alone without having access to GS.

In our analysis of gynaecologic specimens, eight out of ten discrepancies belong to overcalling or under calling of dysplasia in cervical biopsies/large loop excision of transformation zone of cervix specimens. In a study on the validation of WSI in the primary diagnosis of gynecological tumors, eight out of nine discrepancies (88.9%) observed in the study were related to undercall of High grade squamous intraepithelial lesion (H-SIL-CIN2 and CIN3) either as Low grade squamous intraepithelial lesions (L-SIL-CIN1) or reactive changes or negative.<sup>[19]</sup> p16 IHC has also been used to distinguish between H-SIL (CIN2 and CIN3), where it was strongly expressed, and low-grade lesions L-SIL (CIN1), where there was no p16 IHC expression.<sup>[19]</sup> In our study, p16 and Ki67

IHC, was used to distinguish between CIN 1 (low grade) and CIN 2 (high-grade lesions).

Two basic pathologist's errors of omission were observed in our study. Both cases were first, of missed neutrophils in placental membranes in a case of moderate chorioamnionitis and second, of a focus of endometriosis in Tubo-ovarian tissues on DP, which on follow up were as a result of nonsystematic scanning on the digital platform, at both low and medium powers by the reporting pathologists.

There was a case of misidentification error in which the reporting pathologist failed to pay due attention to report entry and incorrectly entered a wrong diagnosis on the DP data sheet, which was promptly picked up at first review, corrected, and feedback was given to the reporting pathologist.

There are reports in the literature that lymphoid and lymphoproliferative disorders are difficult to report on WSI/DP.<sup>[26,27]</sup> This study also found lymphoid disorders challenging to report on DP. One gastric biopsy reported as having a reactive lymphoid tissue on GS was reported on DP by a pathologist with special interest in lymphomas who concluded that the biopsy was suggestive of marginal zone lymphoma, a diagnosis that was upheld by the referral center. In a second case, what appeared to be an increase in the number of atypical lymphocytes in a nasal polyp, and interpreted as nonblastic T-cell non-Hodgkin's lymphoma on DP by a nonlymphoma expert, was correctly diagnosed as a reactive condition on GS by another pathologist with interest in lymphomas.

Both cases were referred to the All Wales Lymphoma Panel where our Ground Truth for each of the two cases was upheld. There was a report in the literature of a case of mixed cellularity Hodgkin lymphoma missed on WSI, later reported as viral lymphadenitis/peripheral T-cell lymphoma, which led the authors to caution would be users of WSI when interpreting lymphoid lesions with inflammatory conditions.<sup>[26]</sup>

The 86 clinically significant variances/discrepancies reported in our study were higher than the 21 reported by Snead *et al.*, 2016,<sup>[1]</sup> who had a similar study sample size. The differences in the number of clinically significant differences reported by both groups might be mainly due to the differences in study design. In our multi-center study, cases were assigned randomly to general pathologists, whereas only experts with subspecialty interests reported the cases in the study by Snead *et al.* Their study was also notably carried out in a single Tertiary University Hospital.<sup>[1]</sup>

In QC in Histopathology, the numbers of layers of processes introduced in the laboratory workflow systems, such as scanning, allocating scanned slides electronically to pathologists and reporting on the digital platform are hypothesized to be directly proportional to the increase in the number of potential errors that could occur in the system. A "DP service management" in departments embarking on running diagnostic services with DP has been recently advocated, which should have oversight for technical



issues, training, governance, and accreditation regarding DP, to ensure viable and sustainable diagnostic DP service management.<sup>[28]</sup>

## CONCLUSIONS

Our study showed a high degree of concordance of DP with conventional GS although we were unable to demonstrate non-inferiority to the criteria set. Cautious approach should be taken to interpret atypia/dysplasia with DP in Skin, GIT, and cervical biopsies as threshold changes slightly with DP.

Expertise in the field and further specialty specific DP training to assess dysplasia is recommended to avoid the error.

At least in the beginning of routine clinical reporting, the pathologists should have an easy access to review the GSs if the diagnostic difficulty arises, particularly in assessing the atypical low-grade foci in prostatic biopsies, suspicious cases of melanomas and atypical lymphomatous processes on DP.

The study also observes that creating a consensus diagnosis on certain lesions among wide range of pathologists in their variable experiences working in different organizations is quite challenging in particularly around benign lesions.

Studies such as ours, are reassuring to the pathology community in the UK and elsewhere, that DP has a unique role in future workforce management plans with its ability to facilitate work carried out in merged departments with the scanners located at a central hub such as in this study set up.<sup>[26]</sup> The advantages of DP in remote reporting could also be used for working from home in the wake of the CoVID 19 pandemic.

We also support a call for more training for users of DP, with focused continuous professional development activities. It would also be prudent for the DP community in the UK and other countries to ensure that the departments of Cellular Pathology seeking to roll out DP are accredited, such as by the International Organisation for Standardisation in the UK (ISO).

In summary, the pattern emerging in this study is that an individual pathologists' background knowledge in the interpretation of a given specimen is more important than the platform on which the report is made, be it DP or GS. Strict adherence to diagnostic criteria by the reporting pathologists always eliminates errors. Users of DP are urged to defer diagnosis to GS if unsure or in doubt of diagnosis.

Our study has been able to verify that the use of DP in routine histological diagnosis produces similar results to reporting on GS while validating the participating pathologists in the use of DP for diagnostic purposes.

## Acknowledgments

We are grateful for the support of the Welsh Government, Cardiff for the Efficiency through Technology Research Grant to the departments of Cellular Pathology in all the Health Boards across Wales, which ensured the completion of this project. The technical support of Sion Davies who scanned

the slides and of Dewi Edwards for Information Technology support are thankfully acknowledged. We are greatly appreciate support of Jane Fitzpatrick, Director of Strategic Programmes, NHS Wales Health Collaborative and Senior Responsible Owner, NHS Wales Digital Cellular Pathology Programme and Melanie Barker, Senior Programme Manager, Diagnostics Services Programme and NHS Wales Digital Cellular Pathology Programme. We also acknowledge the critical review of the manuscript by Dr. Katharina von Loga of the Institute of Cancer Research.

## Financial support and sponsorship

Nil.

## Conflicts of interest

There are no conflicts of interest.

## REFERENCES

1. Snead DR, Tsang YW, Meskiri A, Kimani PK, Crossman R, Rajpoot NM, *et al.* Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology* 2016;68:1063-72.
2. Goacher E, Randell R, Williams B, Treanor D. The diagnostic concordance of whole slide imaging and light microscopy: A systematic review. *Arch Pathol Lab Med* 2017;141:151-61.
3. Fallon MA, Wilbur DC, Prasad M. Ovarian frozen section diagnosis: Use of whole-slide imaging shows excellent correlation between virtual slide and original interpretations in a large series of cases. *Arch Pathol Lab Med* 2011;135:372-8.
4. Thorstenson S, Molin J, Lundström C. Implementation of large-scale routine diagnostics using whole slide imaging in Sweden: Digital pathology experiences 2006-2013. *J Pathol Inform* 2014;5:14.
5. Hanna MG, Pantanowitz L, Evans AJ. Overview of contemporary guidelines in digital pathology: What is available in 2015 and what still needs to be addressed? *J Clin Pathol* 2015;68:499-505.
6. Weinstein RS, Descour MR, Liang C, Bhattacharyya AK, Graham AR, Davis JR, *et al.* Telepathology overview: From concept to implementation. *Hum Pathol* 2001;32:1283-99.
7. Gilbertson JR, Ho J, Anthony L, Jukic DM, Yagi Y, Parwani AV. Primary histologic diagnosis using automated whole slide imaging: A validation study. *BMC Clin Pathol* 2006;6:4.
8. Wilbur DC, Madi K, Colvin RB, Duncan LM, Faquin WC, Ferry JA, *et al.* Whole-slide imaging digital pathology as a platform for teleconsultation: A pilot study using paired subspecialist correlations. *Arch Pathol Lab Med* 2009;133:1949-53.
9. Bauer TW, Schoenfield L, Slaw RJ, Yeran L, Sun Z, Henricks WH. Validation of whole slide imaging for primary diagnosis in surgical pathology. *Arch Pathol Lab Med* 2013;137:518-24.
10. Campbell WS, Lele SM, West WW, Lazenby AJ, Smith LM, Hinrichs SH. Concordance between whole-slide imaging and light microscopy for routine surgical pathology. *Hum Pathol* 2012;43:1739-44.
11. Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, *et al.* Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med* 2013;137:1710-22.
12. Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, *et al.* Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: A multicenter blinded randomized noninferiority study of 1992 cases (Pivotal Study). *Am J Surg Pathol* 2018;42:39-52.
13. Raab SS, Grzybicki DM, Mahood LK, Parwani AV, Kuan SF, Rao UN. Effectiveness of random and focused review in detecting surgical pathology error. *Am J Clin Pathol* 2008;130:905-12.
14. Griffin J, Treanor D. Digital pathology in clinical use: Where are we now and what is holding us back? *Histopathology* 2017;70:134-45.
15. Raab SS, Nakhleh RE, Ruby SG. Patient safety in anatomic pathology:

- Measuring discrepancy frequencies and causes. *Arch Pathol Lab Med* 2005;129:459-66.
16. Tsung JS. Institutional pathology consultation. *Am J Surg Pathol* 2004;28:399-02.
  17. Frable WJ. Surgical pathology Second reviews, institutional reviews, audits, and correlations: What's out there? Error or diagnostic variation? *Arch Pathol Lab Med* 2006;130:620-5.
  18. Ordi J, Castillo P, Saco A, Del Pino M, Ordi O, Rodríguez-Carunchio L, *et al.* Validation of whole slide imaging in the primary diagnosis of gynaecological pathology in a University Hospital. *J Clin Pathol* 2015;68:33-9.
  19. Al-Janabi S, Huisman A, Vink A, Leguit RJ, Offerhaus GJ, Ten Kate FJ, *et al.* Whole slide images for primary diagnostics in dermatopathology: A feasibility study. *J Clin Pathol* 2012;65:152-8.
  20. Elmore J, Barnhill R, Elder D, Gary L, Margaret PS, Lisa R, *et al.* Pathologists' diagnosis of invasive melanoma and melanocytic Proliferations. *Br Med J* 2017;357:j2813.
  21. Williams BJ, DaCosta P, Goacher E, Treanor D. A systematic analysis of discordant diagnoses in digital pathology compared with light microscopy. *Arch Pathol Lab Med* 2017;141:1712-8.
  22. Gui D, Cortina G, Naini B, Hart S, Gerney G, Dawson D, *et al.* Diagnosis of dysplasia in upper gastro-intestinal tract biopsies through digital microscopy. *J Pathol Inform* 2012;3:27.
  23. Van der Kwast TH, Evans A, Lockwood G, Tkachuk D, Bostwick DG, Epstein JI, *et al.* Variability in diagnostic opinion among pathologists for single small atypical foci in prostate biopsies. *Am J Surg Pathol* 2010;34:169-77.
  24. Rodriguez-Urrego PA, Cronin AM, Al-Ahmadie HA, Gopalan A, Tickoo SK, Reuter VE, *et al.* Interobserver and intraobserver reproducibility in digital and routine microscopic assessment of prostate needle biopsies. *Hum Pathol* 2011;42:68-74.
  25. Ozluk Y, Blanco PL, Mengel M, Solez K, Halloran PF, Sis B. Superiority of virtual microscopy versus light microscopy in transplantation pathology. *Clin Transplant* 2012;26:336-44.
  26. Cheng CL, Tan PH. Digital pathology in the diagnostic setting: Beyond technology into best practice and service management. *J Clin Pathol* 2017;70:454-7.
  27. Williams BJ, Bottoms D, Treanor D. Future-proofing pathology: The case for clinical adoption of digital pathology. *J Clin Pathol* 2017;70:1010-8.
  28. Borowsky AD, Glassy EF, Wallace WD, Kallichanda NS, Behling CA, Miller DV, *et al.* Digital whole slide imaging compared with light microscopy for primary diagnosis in surgical pathology: A multicenter, double-blinded, randomized study of 2045 cases. *Arch Pathol Lab Med* 2020,144(10):1245-53.